

A

Very interesting -
You might do me by way
of comparing your
project to that of
other -

Fact Schmact and Science Fiction

Patrick Grim

Boston

December, 1972

Fact Schmact and Science Fiction

I. Introduction

All our problems are getting married.' Soon there will be no bachelor queries or spinster questions left, and the philosophical halls will echo with the pitter-patter of little problems related as brother and sister.

A prime example, of course, is the so-called 'problem of counterfactuals', which is in fact a 'nest of problems stacked inside each other like Chinese boxes', 'an unholy alliance', or 'a family of related questions' - anything, in fact, but a monolithic whole to which a yes or no answer is appropriate. It is this conglomeration of worries which I hope to some extent to clarify and to a significantly lesser extent to abate. My first effort will be to clarify the problems involved, their importance, and the nature of their relation. A second section involves in proposal a set of criteria for distinguishing between true and false counterfactuals.

II. The Problem

There are basically two ways of systematically viewing the 'problem of counterfactuals', and from both vantages it appears as a snarled mess.

The first is on the basis of its relation to other important issues. Goodman writes:

"Indeed, if we lack the means for interpreting counterfactual conditionals, we can hardly claim to have any adequate philosophy of science. A satisfactory definition of scientific law, a

satisfactory theory of confirmation or of dispositional terms ... would solve a large part of the problem of counterfactuals. Conversely, a solution to the problem of counterfactuals would give us the answer to critical questions about law, confirmation, and the meaning of potentiality."¹

Chisholm notes:

"As we shall see, the philosophical problems which this question involves are fundamental to metaphysics, epistemology, and the general philosophy of science."²

A large number of paths (to change metaphor in mid-essay) converge in counterfactuals. An apparently crucial difference between statements we normally consider 'causal' or 'natural' laws and ordinary general statements is that the former 'support' counterfactuals and the latter do not. A definitive characteristic of 'dispositional predicates' is that they may be rephrased in terms of counterfactuals. When asked what we mean by saying 'it is possible that ... ', we tend to give an answer which is counterfactual in form. Laws and possibility and dispositional terms are all tied together. And the knot that binds them is the troublesome counterfactual.

The second way of viewing 'the problem of counterfactuals' is that available when we ask for a characterization not of the importance but of the nature of such a grammatical form. The 'nature' question has been sketched as falling into three parts:

The 'logical' problem: "This is the task of describing the formal properties of the conditional function: a function, usually represented in English by the words 'if ... then', taking ordered pairs of propositions into propositions."³

The 'pragmatic' problem: "The formal properties of the conditional function, together with all of the facts, may not be sufficient for determining the truth value of a counterfactual; that is, different truth valuations of conditional statements may be consistent with a single valuation of all non-conditional statements. The task set by the problem is to find and defend criteria for choosing among these different valuations."⁴

The 'epistemological' problem: "How are conditionals which are both empirical and contrary-to-fact possible at all? How do we learn about possible worlds, and where are the facts (or counterfactuals) which make counterfactuals true?"⁵

The question of the nature of counterfactuals, like that of their importance, is many-sided. It includes both questions of translatable form (what non-problematic [read 'indicative'] grammatical construction is equivalent to counterfactuals? What formal semantic is possible as an expression of counterfactuals? Is there an equivalent to counterfactuals expressible in canonical logical form?) and of linguistic function (how does one decide between a counterfactual and its contrary - that with the same antecedent and contradictory consequent)? The third or 'epistemological' question is largely derivative from the other two; if we are to accept a Leibnizian account of counterfactuals in terms of non-actual worlds, then how are we to outline confirmation (which, if it includes counterfactuals, must on such an account include confirmed non-actuals)? The 'logical' and 'pragmatic' questions are simple queries. The 'epistemological' question is based on, and may be used as an argument against, a possible answer to the first two.

The 'logical' and 'pragmatic' questions of the nature of counterfactuals are methodologically tied; an answer to the first might well give us a clue to the answer to the second. But it is the second that is primary, for a formal answer (such as Stalnaker's⁶) may throw as little light on the function of counterfactuals in ordinary usage as 'were ... would' throws on our understanding of dispositional terms.

Thus although the 'logical' question may indeed throw light on the 'pragmatic' question, it need not. The simple existence of a formal system reflecting the structure of ordinary counterfactuals may tell us nothing of why we think some are true and some false; and it is that which we really want to know.

We have thus perhaps indicated a central snarl of the tangled skein. The primary problem, as Goodman outlines it,

"...is to define the circumstances under which a given counterfactual holds while the opposing conditional with the contradictory consequent fails to hold. And this criterion of truth must be set up in the face of the fact that a counterfactual by its nature can never be subjected to any direct empirical test by realizing its antecedent."^{7a}

In what follows I hope to make at least some shaky first steps towards supplying such a definition. The goal is a set of criteria, according more or less with our intuitions, and supplying a method of choosing between true and false counterfactuals on the basis of nothing more celestial or surreal than a mundane knowledge of the facts.

III. All that glitters is not gold and all that functions is not

fact. In what follows I hope to outline the nature of what I shall term 'restrictive fictions'; those fictions, roughly, that are not stranger than fact. My thesis is that an understanding of 'restrictive fictions' contributes to our understanding of counterfactuals.

A. A necessary first step in this direction is the isolation of a number of troublesome cases which any adequate set of criteria must satisfy. A number of other obstacles may arise in the course of our exposition, but these may serve as foundations for a first attempt.

1. Not all unimmersed things are soluble; it is not true of any arbitrarily chosen thing that 'were it immersed, it would dissolve'.

2. A counterfactual may be true although its antecedent is never and could never be satisfied; although yesterday's butter was consumed with this morning's toast, and is thus forever beyond the reach of empirical test, it seems acceptable to say 'were yesterday's butter heated to 150 degrees farenheit, it would have melted.'

3. We may maintain both of the following; All properly constructed matches, struck in enough oxygen, light. All the coins in my pocket are silver. Yet we distinguish between: 'If that match were struck, it would light', and 'If that coin were in my pocket it would be silver.'

4. Any individual thing is a member of a number of non-equivalent classes. These classes, moreover, may be highly consistent with contradictory properties. Thus it is true that a high number of things which are either this pencil or glass snap in two when pressure

Key under

is applied. It is also true that a high number of things which are either this pencil or powder do not snap when pressure is applied. But we distinguish between: 'Were pressure applied to this pencil, it would snap' and 'Were pressure applied to 'this pencil, it would not snap'.

B. What kind of criteria, appealing to nothing more outrageous than the (indicative) facts, may be constructed to deal with these? Let us first border our immediate concern a bit, and take as our goal an outline of singular counterfactuals whose antecedent and consequent have the same grammatical subject. A first proposal, using terms to which the reader may not as yet have been formally introduced, follows.

Allow me to present to you ...

A 'name' is a word used as a noun which designates a set of one individual. Thus "Joe" is not a name, but "Joe Emelius Xavier De Bonzo", I believe, is.

A 'type' is any term which designates, or whose extension includes (in the case of predicates) one or more individuals.

A 'faction' is a true set of sentences, which is the same as a set of true sentences.

A 'fiction' is a false set of sentences, which is not always the same as a set of false sentences.

A set p is perfectly 'homogeneous' viz a viz a property q if and only if all members of p are q (taking q as a 'type' functioning as a predicate, the set of p is a subset of q). A set p is more homogeneous viz a viz a property q than is a set r viz a viz a property s if and only if a higher percentage of p 's members are q than r 's members are s (taking q and s as predicate 'types', the

ratio of those members of p which are also members of q over (/) all members of p is closer to 1 than the ratio of those members of r which are also members of s over all members of r.)

A set p is perfectly 'homogeneous' viz a viz two properties q and r if and only if all p's which are q are also r (the intersection of p and q is a subset of r). A set p is more homogeneous viz a viz q and r than is a set s viz a viz t and u if and only if a higher percentage of the p's which are q's are also r's than the s's which are t's are also u's (if and only if the ratio of those members of p which are members of the intersection of q and r over (/) those members of the intersection of p and q is closer 1 than is $t/s \wedge u / t/s$).

The following, then, constitutes a first proposal:

A counterfactual condition of the form 'If x were y, x would z', where 'x' is a name, is true if and only if, for the indicative p corresponding to the counterfactual ('x is y and x is z'):

(i) There exists a name faction (a set of true sentences whose grammatical subjects designate classes of one member) such that when p is taken in conjunction with it, it becomes a name fiction (a false set of sentences whose grammatical subjects designate classes of one member).

(ii) The name fiction thus produced is amenable to type substitutions for x alone or for x and (y and/or z) such that that substitution creates a type faction (substitution for y and/or z will obviously be in terms of predicate types).

(A set of type sentences is true - is a type faction - if and only if for some substitution of individual members for the 'types' to which they correspond, a set of true name sentences is produced.)

*Don't you want
be name
K is not in
subject
take other
the
subset*

(iii) x, y, and z are members of the sets substituted for them as types.

(iv) The set a substituted for x in (ii) is more homogeneous viz a viz y and z (or their substituted types) than is any alternative substitutable set (set of which x is a member) viz a viz y and not-z.

(v) The set a to which x is referred must not be such that there exists a subset b of a such that 1/ x is a member of b, 2/ the 'residue' c of b in a (the set of members of a which are not members of b), of which x is not a member, is or includes the intersection of a and y, and 3/ that subset is more homogeneous viz a viz y (or some type substitution for y) and not-z than is a viz a viz y or that substitution (whichever is used) and z.

C. The exact application of our schema should become clearer by way of consideration of the initial problems we have set ourselves. The intuition behind (iv) in particular, however, should be obvious. We often at first accept a counterfactual of the form 'this radiator, if frozen, would explode'. On further examination, however, we note that the radiator is not filled with water, and we retract our original ascent to the condition. (v) is similarly based on a common intuition; namely that there is something better about justifying the counterfactual 'this would break if frozen' by pointing out that the object in question is a radiator than by assigning it to some arbitrary class. The difficulties we face are primarily those of making these intuitions explicit. Let us turn to the initial problems:

1) Not all unimmersed things are soluble. This is in accord with our criteria, for it is true of some (non-immersed) individual

and in you

that 'if it were immersed, it would dissolve' only if that individual is a member of a set some member of which, at some time, is immersed and dissolved. *here is your first trouble - the counterfactual might be true even if nothing were ever melted*

2) Although yesterday's butter is now where mortal experiments can no longer harm it, it is nonetheless true that it would have melted had it been heated. It belongs to the class 'butter', and 'butter is heated and melts' is a fact. ~~but it is not a fact that butter is heated and melts~~

3) Consider the following:

(i) Everyone in this room speaks English.

(ii) If Khrushchev is in this room, he speaks English.

(iii) If Khrushchev were in this room, he would speak English.

Although (ii) - not a counterfactual, but more properly speaking a "don't know conditional" - is legitimate in light of (i), (iii) is not. The reason for this becomes a bit clearer when we also consider an apparently legitimate (iv):

(iv) If Khrushchev were in this room, not everyone in this room would speak English.

This, again, is not the type of counterfactual we have been examining; the grammatical subjects of its antecedent and consequent are not the same. It nonetheless throws light on the rejection of (iii) by our criteria. Our intuitions are that the reason that (iii) is illegitimate is that (iv), involving the same antecedent and a consequent denying (i), which is in some sense the 'ground' of (iii), is legitimate. But how are we to avoid Goodman's conclusion that knowledge of the truth or falsity of a counterfactual presupposes knowledge of the truth or falsity of another, ad infinitum?^{7b}

A solution, embodied in our fifth criterion, centers on the observation that it is the same antecedent involved in (iii) and (iv).

An infinite regression is thus avoidable by building into our schema basic conditions concerning the falsehood of (iv), but expressed in terms of formal properties of x and a viz a viz y and z alone.

This is, as we shall see, precisely what (v) supplies.

First, however, a minor digression may be helpful. 'Were Khrushchev in this room, he would speak English' may be used in two quite different ways. The more 'legitimate' is often signalled by a preference for 'If Khrushchev were one of the people in this room, he would speak English', or (ii) simpliciter. If the people in this room are Joe, Mike, and Fred, then 'Khrushchev' is referred to the class of people which are Joe, Mike, or Fred. As that class is homogeneous viz a viz speaking English, the counterfactual is legitimate.

In the usual sense of the counterfactual, however, this is not the case. (iii), as usually interpreted, refers Khrushchev to the class of people in general. This class, it must be noted, is homogeneous viz a viz being in this room and speaking English. But it is obvious that it is 'Joe is in this room and speaks English', 'Mike is in this room and speaks English', and 'Fred is in this room and speaks English' which are the name factions which make 'People who are in this room speak English' a faction. Joe, Mike, and Fred constitute the intersection of a (all people) and y (things in this room). As Khrushchev is not a member of the set of Joe, Mike, and Fred, he may be assigned to a 'residue set' of a (all people) ; namely 'Russians' or 'Boris, Natasha, Illya, and Khrushchev'. We may further substitute for y (things in this room) a more general type (things in a place). The result is that

the set 'Russians' to which Khrushchev is assigned is more homogeneous viz a viz being in a place and not speaking English than is the class of all people viz a viz being in a place and speaking English. It thus appears that our fifth criterion effectively blocks the legitimacy of (iii), though saving that of (ii).

A first move towards understanding (iii), then, is to sort out its ambiguities. Once that is done, our fifth criterion can provide a legitimate decision procedure, corresponding to Mackie's suggestion⁸, but retaining the notion of counterfactuals as in some sense true or false (rather than simply "condensed arguments").

The fifth criterion is also effective against 'Were this penny in my pocket, it would be silver' ('All coins in my pocket are silver'). 'This penny' (x) is assigned to the class of all coins (a), which is highly homogeneous viz a viz being in my pocket (y) and being silver (z). 'This penny', however, is 1/ a member of a subset of a, namely 'pennies', 2/ the residue of which in a includes the intersection of a and y. The set of 'pennies' is furthermore 3/ more homogeneous viz a viz being in a place (a type substitution for y) and not being silver than is the class of all coins (a) viz a viz being in a place and being silver. It is thus (v) once again that blocks the legitimacy of the counterfactual in question. (An ambiguity similar to that of (iii) may also be noted in this case; 'were that penny one of the coins which are in my pocket, it would be silver', under the usual interpretation, appears to be legitimate).

4) Here, as in the case of (3), it is our fifth criterion that is the decisive factor. Consider the apparently legitimate counterfactual:

(i) Were that pencil subjected to pressure (of a certain type), it would break.

and its apparently illegitimate contrary;

(ii) Were that pencil subjected to pressure (of a certain type) it would not break.

By assigning the pencil in question to the class of everything that is that pencil or powder, it appears that we may illegitimately justify (ii) on the grounds that 'All things that are pencil x or powder and are subjected to pressure (of a certain type) do not break'.

Our fifth criterion, however, is once again effective against such a move. We may treat the class of assignment (a) as 'all things', the first property (y) as 'thing which are that pencil or powder and are subjected to pressure', and the second property (z) as 'thing which break'. But now it is obvious, in violation of our fifth criterion, that 1/ there exists a subset of a, namely 'all pencils', of which x is a member, 2/ the 'residue' of that set includes the intersection of all things (a) and things which are that pencil or powder and are subjected to pressure (y) [as x is not subjected to pressure, x is not a member of that intersection], and 3/ the set of 'all pencils' (b) is more homogeneous viz a viz a type substitution for y - all things subjected to pressure - and z than is the class of all things viz a viz subjection to pressure (the type substitution for y) and not breaking.

It thus appears that our criteria, as they stand, are effective against the initial problems we set ourselves. A number of further virtues of our schema should also be noted:

5) There are two problems with respect to empty sets and counter-

factuals. The first, noted by Goodman⁹, concerns assignment of the individual x to an empty set. Our schema is such, however, that this assignment is impossible ('assignment' in our sense must be to a class of which x is a member, and thus cannot be empty), and thus the problem does not arise. Goodman's expansion of that problem in terms of disjunctive sets, one part of which is empty, may be treated along the lines of (3) and (4).

But what of counterfactuals such as the following:

(1) Were Pegasus real, Pegasus would fly.

This appears to be at least a candidate for legitimacy, yet Pegasus is a mythical animal, and the faction of the form 'All mythical animals which are real ...' seems to be necessarily empty.

Our schema, however, does allow an assignment of types such that (1) is seen to be legitimate. We need not refer Pegasus to the class of mythical animals; we may rather refer Pegasus to the class of animals, which under at least one interpretation includes the class of mythical animals. Taking as a first predicate 'being mentioned in stories as having wings' and as a second predicate 'flying', we may construct the faction 'Animals which are mentioned in stories as having wings fly'. It is via a substitution of this sort that (1) in particular and counterfactuals whose grammatical subjects are non-existent objects in general may be treated.

6) We noted, in a first section, the close connection between (among other things) counterfactuals and 'causal' or 'natural' laws. If our proposal is at least effective within a limited domain, what sketch of lawlikeness does it offer us?

It should be noted with respect to our schema that the difference

between 'laws of nature' and 'accidental generalizations' fails to be one of strict and formal structure. The difference is rather one of degree than of kind, for the failure of 'accidental generalizations' to "support" counterfactuals is accountable simply in terms of more homogeneous classes of assignment. The notion of 'law', on our outline, becomes a relative notion, precisely as confirmation of counterfactuals becomes relative to the homogeneity of assignable types.

7) Science as a whole, our criteria suggest, is not simply and completely a matter of fact. Restrictive fictions, the type of false set of sentences which are not stranger than fact, are essential to counterfactuals and thus to a major portion of the scientific enterprise. To master counterfactual locutions is to master a type of story-telling which relies essentially on ambiguity and general reference. The scientific world is the totality of facts and restrictive fictions, not of things.

8) It should finally be remembered that ours are criteria for truth of counterfactuals of a stringently limited type. Although the foundations of a correlate treatment of counterfactuals the grammatical subjects of whose antecedents and consequents are not the same was indicated in (3), our treatment has been explicitly confined to counterfactuals whose grammatical subjects are individual and identical in antecedent and consequent clauses. Other types of counterfactuals, as well as what Goodman isolates as 'semifactuals', 'counteridenticals', and the like, remain to be accounted for.

But to the extent that our criteria present a first step towards the understanding of a limited type of counterfactuals, they also present a first step towards the understanding of these related locutions.

IV. The Last Bow

Unfortunately, a number of more troublesome problems remain. Although a first step may have been made within a limited realm, that accomplishment is qualified by the surrounding territory of untreated questions. And those limiting domains can no longer be treated as bachelor questions or spinster queries, extended condescending invitations to tea at our philosophical leisure.

It must be noted in this context that the 'types' and 'predicates' of our criteria are, as they stand, outlined in terms of classes of individuals. But if we have a finite number of individuals, we have as well a finite number of non-overlapping 'predicates' or 'types' in this sense. In a world of three individuals, and predicates in this sense alone, our types are $\{a\}$, $\{b\}$, $\{c\}$, $\{a b\}$, $\{a c\}$, $\{b c\}$, and $\{a b c\}$. But in such a case no counterfactuals are possible at all, for there are classes to which any x is assignable, and which are equally homogeneous viz a viz contrary properties.

We may, however, simply gulp hard and admit that counterfactuals are only possible in a world with disproportionate assignment of individuals to classes. But a second limiting problem remains.

If we admit to such disproportionate assignment in natural languages, how are we to justify the assignments as they stand? We do class green things together. But why not grue ...

All our problems are getting married.

Notes

- ¹Nelson Goodman, "The Problem of Counterfactual Conditionals", in Fact Fiction and Forecast.
- ²Roderick Chisholm, "The Contrary-to-Fact Conditional", in Herbert Feigl and Wilfrid Sellars, ed., Readings in Philosophical Analysis.
- ³Robert C. Stalnaker, "A Theory of Conditionals", in Cornman, et. al, Studies in Logical Theory, American Philosophical Quarterly Monograph No. 2.
- ⁴Ibid.
- ⁵Ibid.
- ⁶Stalnaker, op. cit.
- ⁷^aGoodman, op. cit.
- ⁷^bIbid.
- ⁸J. L. Mackie, "Counterfactuals and Causal Laws", in R. J. Butler, ed., Analytical Philosophy, first series.
- ⁹Goodman, op. cit.